Author Names & Affiliations

- Randolph Moses - The Ohio State University

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

multiple

Title of Submission

Ohio State University Institutional Response

Abstract (maximum ~200 words).

At a large public institution, cyberinfrastructure needs include multiple scales and cover a large range of disciplinary and cross-disciplinary research areas. A continuing challenge remains how to accommodate growing demand and how to provide computational resources and data science expertise to a growing user community, and to support both the physical and human expertise infrastructure to researchers, especially early-career researchers. In particular, the needs for pervasive small- and medium-scale cyberinfrastructure are more difficult to meet with established practices and current funding models. Research groups need access to shared services specializing in experimental design, data management, data analytics, learning algorithms, and similar topics. Mechanisms to support permanent research scientists who can provide both expertise and training of modern cyberinfrastructure tools to disciplinary researchers, especially early-career researchers, would benefit the research community. NSF could provide benefit to researchers by providing leadership within the federal government on the special needs of researchers, and by considering appropriate exemptions or broad definitions to research data security in emerging implementations and applications.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

The Ohio State University (OSU) is a major public land grant institution with 66,000 students, 45,000 employees, a large Medical Center, and annual research expenditures exceeding $900 million. Our faculty, staff, and students are engaged in virtually every category of research that NSF supports. We use cyberinfrastructure on every scale, from the largest simulations for the physical and life sciences to tiny sensor/processors in engineering, agriculture, and medicine. Major interdisciplinary faculty hiring and research initiatives will continue to grow these efforts in cross-cutting thematic areas; data science will play a major role in all of these.

A continuing challenge remains how to provide cyber infrastructure support to engage our research teams toward both disciplinary data science research questions and trans disciplinary research that engages teams of both data science researchers and disciplinary researchers. An example of this would be teams developing production and use models for sustainable agriculture in geographic regions, in which researchers in agricultural production, water resources, economics, and social science provide create complex multi-disciplinary models that support scientific predictions, and data scientists engage in developing efficient ways to implement these models.

OSU researchers have excellent access to high-performance computing and networking through the Ohio Supercomputer Center (OSC) and the Ohio Academic Research Network (OARnet), both located on OSU's campus and generously supported by the State of Ohio for many years. OSC recently installed its $11 million Owens cluster (22,000 Broadwell cores, 160 Nvidia P100 GPUs), available to Ohio PIs at no charge. OARnet's 100 gigabit/second backbone connects Ohio campuses to Internet2, to NSF's XSEDE infrastructure, and to DOE's ESnet-connected facilities. These strong local facilities meet the complete needs of many OSU researchers, while providing a convenient development environment for more demanding users as they progress to large-scale NSF and DOE facilities. The concerns of OSC and OARnet closely parallel those described in the National Academies' study "Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering 2017-2020" (2016). While this excellent infrastructure exists, a continuing challenge remains how to accommodate growing demand and how to provide computational resources and data science expertise to a growing user community, and to support both the physical and human expertise infrastructure to researchers, especially early-career researchers.

OSU's needs for pervasive small- and medium-scale cyberinfrastructure are more difficult to meet with established practices and current funding models.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

While Ohio State's researchers benefit from steady improvements in computer hardware, software, networking, and storage, their productivity is continually challenged. This challenge takes many forms:

• Modern cyberinfrastructure tools are used throughout our various research disciplines, but few researchers possess both the technical skills and the domain knowledge to use them effectively. NSF would benefit the research community by providing mechanisms to support permanent research scientists to fulfill this role. Current solutions are less efficient: many times these 'experts' PhD students who learn the needed skills, then provide those skills to their local research group. Graduate students and postdocs are transient, so there is significant turnover in these functions, and it is difficult to maintain corporate knowledge. Universities often lack resources to provide such specialized functions as part of their research overhead infrastructure. Without dedicated, long-term experts, many research groups make inefficient use of cyberinfrastructure.

• Modern cyberinfrastructure is often complex and difficult to use. While newer tools and packaged software have improved the situation, other forces have aggravated it, including: increased data volume and complexity; heterogeneous and cloud computing; and the proliferation of distributed sensors and actuators. Better "portals" are needed to simplify user's access to complex systems. The Ohio Supercomputer Center is currently building a toolkit for portals with NSF support. While a good start, more effort is needed in this area.

• Most universities, including Ohio State, seek better ways to train researchers, especially their graduate students, in the use of modern cyberinfrastructure. Universities are experimenting with service courses, majors, minors, and graduate specializations, and NSF is funding many of these experiments. Efforts to broadly disseminate the results of these experiments and to assemble the produced training materials for the research community would be welcome.

• Research groups need access to shared services specializing in experimental design, data management, data analytics, learning algorithms, and similar topics. Most groups do not need such services frequently enough to justify their own experts, but need occasional advice and support to avoid problems. Universities have long maintained shared services around specialized instruments and laboratory animals, and now need them for cyberinfrastructure. NSF support to incentivize and facilitate shared services would benefit our researchers.

• Recent requirements for long-term public access to research data remain challenging in definition and implementation. More than 1,500 research data repositories now exist, sponsored by universities, professional societies, funding agencies, and others. Some repositories handle many varieties and scales of data, while others are more restrictive, and some handle only the metadata of data stored elsewhere. OSU is partnering in an NSF Big Data Regional Innovation Hub and is on the Board of SHARE, a large metadata repository. Questions remain about long-term sustainability of these repositories, and infrastructure or metadata to ensure that data products are valuable for the long term. Sustainability will require ongoing reinvestment as data formats and tools evolve. The costs of maintaining data access infrastructure are unknown, and it is not clear how the national research community will fund them. NSF would benefit the community by continuing to provide leadership in defining and refining public access needs, and by providing resources to support those needs.

• The open-source software community is producing excellent code. It is less well equipped, to provide maintenance. Some of this code is extremely valuable to relatively small groups of researchers in a community, and it does not have the scale to provide enough resources for maintenance. In addition, work is needed to integrate these tools into production pipelines and to adapt them for new computer architectures. Some mechanism is needed to fund these essential services, while retaining the low-cost benefits of open-source, would benefit the research communities. In each of these areas, the chief problem is sustainability. Researchers receive (episodic) grant funding, including overhead funds to cover universities' normal administrative and operating costs. Ohio State as an institution struggles with the questions of how to effectively support our research community to meet these needs.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

Federal requirements and best practices for data security generate an increasing resource burden on universities, in an environment in which administrative cost recovery for federal research funding is capped. Moreover, in many cases these requirements are written for data stored within a traditional data center, with no provision for laboratory computers, mobile computing for field researchers, or the distributed sensors and actuators of the "Internet of Things" that is often found in research communities. NSF could provide benefit to researchers by providing leadership within the federal government on the special needs of researchers, and by considering appropriate exemptions to research data security in emerging implementations and applications.

Consent Statement